**Report**

**Stock Price Prediction using Machine Learning:**

**Abstract**

This project aims to predict the future closing price of Tesla's stock (TSLA) using machine learning techniques. By leveraging historical stock data, including price information (open, high, low, close) and trading volume, we apply feature engineering to enhance the dataset and utilize a Random Forest Regressor model to make predictions. The model is evaluated using common regression metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R²). The results of this analysis show how machine learning models can be used for forecasting financial time series data, providing valuable insights into stock price prediction.

**Introduction**

Stock price prediction is an important task in the financial industry, with applications ranging from algorithmic trading to risk management. The goal of this project is to use historical data to predict Tesla’s (TSLA) stock closing price for future dates. Machine learning models, particularly ensemble methods like Random Forest Regressor, are employed to learn patterns from historical data, such as stock prices and trading volume, and generate forecasts for unseen data. This report covers the dataset used, the methods employed for data manipulation and visualization, the preprocessing steps, model training, and evaluation, and concludes with insights gained from the analysis.

**Dataset Description**

The dataset used in this project contains historical stock data for Tesla (TSLA), including key features such as:

* **Date**: The trading date.
* **Open**: Opening stock price.
* **High**: Highest stock price during the trading day.
* **Low**: Lowest stock price during the trading day.
* **Close**: Closing stock price, which is the target variable.
* **Volume**: Number of shares traded.

The dataset is loaded from a CSV file and indexed by the **Date** column to facilitate time-series analysis.

**Technologies Used**

The following technologies were used in the project:

* **Python**: The primary programming language for data manipulation, model training, and evaluation.
* **Pandas**: A powerful library for data manipulation and analysis, used to load and preprocess the dataset.
* **NumPy**: A library for numerical computing, used for array manipulations and mathematical operations.
* **Matplotlib** and **Seaborn**: Used for data visualization to plot graphs and evaluate the performance of the model.
* **Scikit-learn**: A machine learning library in Python used for implementing and training the Random Forest Regressor model and evaluating its performance.

**Data Preprocessing**

**Handling Missing Values:**

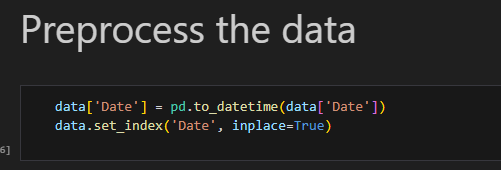
The dataset may contain missing values, especially due to the creation of lag features or rolling means. The missing values are dropped using the following code to ensure clean data for modeling:



**Data Manipulation and Visualization**

**Data Manipulation:**

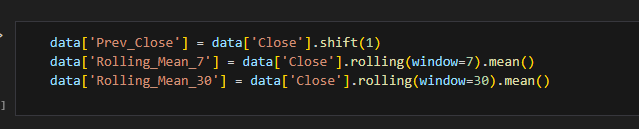
The dataset is loaded into a Pandas DataFrame, and the **Date** column is converted to a datetime format for easier manipulation. The **Date** column is then set as the index, which is essential for time-series analysis.



**Feature Engineering:**

To enhance the model’s ability to make accurate predictions, several new features are created:

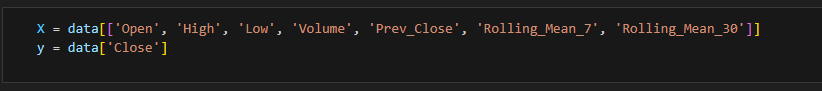
* **Prev\_Close**: The closing price from the previous day.
* **Rolling\_Mean\_7**: The 7-day rolling mean of the closing price, capturing short-term trends.
* **Rolling\_Mean\_30**: The 30-day rolling mean, which identifies long-term trends.



**Data Visualization:**

We visualize the actual vs. predicted closing prices to understand the performance of the model visually. For instance, the following plot compares the actual and predicted closing prices of Tesla stock:

**Splitting the Dataset:**

The data is split into features (**X**) and target (**y**). **X** includes features like Open, High, Low, Volume, Prev\_Close, Rolling\_Mean\_7, and Rolling\_Mean\_30, while **y** represents the target variable, which is the **Close** price.**Model Training and Evaluation**

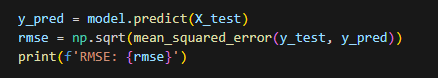
**Training the Model:**

A **Random Forest Regressor** model is trained on the training set (**X\_train**, **y\_train**). Random Forest is a powerful ensemble machine learning algorithm that works well with time-series data for predicting stock prices.

****

**Model Evaluation:**

After training, the model is used to make predictions on the test set. We evaluate the model using **Root Mean Squared Error (RMSE)**, **Mean Absolute Error (MAE)**, and **R² (R-squared)** metrics.



These evaluation metrics help in assessing the model’s performance and the accuracy of predictions.

**Conclusion**

This project demonstrates the use of machine learning, specifically the **Random Forest Regressor**, for stock price prediction. By utilizing historical Tesla stock data and engineering relevant features such as previous close prices and moving averages, the model is able to make predictions on future stock prices.

The model’s performance was evaluated using key metrics such as RMSE, MAE, and R², with visualizations further highlighting the accuracy of the predictions. The results indicate that machine learning models like Random Forest can be effective tools for stock price forecasting, though financial predictions are inherently uncertain and subject to various external factors. This model can be enhanced by using more advanced techniques or incorporating additional external data, such as market sentiment or news data, for better accuracy.